# Mode predictors in nonlinear systems with identities

Giorgio Calzolari and Lorenzo Panattoni

IBM Scientific Center, Pisa, Italy.

September 1988

# MODE PREDICTORS IN NONLINEAR SYSTEMS WITH IDENTITIES

*by* Giorgio CALZOLARI *and* Lorenzo PANATTONI

## ABSTRACT

The shortcomings of predictors obtained with the usual deterministic solution methods in nonlinear systems of stochastic equations have been widely investigated in the literature. Most of the proposed therapies are based on some estimation of the conditional mean of the endogenous variables in the forecast period. This however provides a solution to the problem which does not respect the internal coherency of the model, and in particular does not satisfy nonlinear identities. At the same time, for analogy with univariate skewed distributions, the conditional mean may be expected to lie on the *wrong* side of the deterministic solution, meaning that it moves towards values of the variables where the probability density is lower, rather than towards the most probable values. In a previous study, Brillet, Calzolari and Panattoni (1986) proposed to estimate the *mode* of the joint distribution of the endogenous variables as an alternative optimal predictor. The method proposed in that paper maximized the joint density of a subset of the endogenous variables, corresponding to stochastic equations only (analogously to FIML estimation where, at least conceptually, identities are first substituted into stochastic equations, and then the likelihood is defined and maximized). In this paper, a more general approach is developed, which maintains the identities. The model with identities is viewed as a mapping between the space of the random errors and an hypersurface in the higher dimensional space of the endogenous variables; maximization is performed on such a hypersurface. Experimental results on these two *mode predictors* (and comparisons with the *mean predictors* that are more usually proposed as alternatives to the deterministic predictors) are provided for a macro model of the Italian economy. (*)

(*) A simplified approach to the problem was proposed at the *1986 European Meeting of the Econometric Society* (Budapest) in a joint paper with Jean-Louis Brillet, and experimental results were given for a large scale macroeconomic model of the French economy developed at INSEE. After suggestions and comments received from F.J. Henk Don, Steve G. Hall, Fortunato Pesarin, Frederic P. Sterbenz, and Kenneth F. Wallis, in this paper we tackle the problem from a more general point of view. Mirella Damiani supplied the data and the model experimented with at the end of the paper. We are greatly indebted to all of them. A particular thank is due to Giorgio Fotia, for suggesting us the mathematical framework that underlines the new method. However, responsibility for the contents of this paper remains only with the authors.



## CONTENTS

## 1. INTRODUCTION

Predictors from nonlinear systems of stochastic equations are usually produced with deterministic simulation. The shortcomings of this method have been widely investigated in the literature.

For example, being deterministic solution usually different from the conditional mean of the endogenous variables, model's validation should not be based on the comparison between deterministic solution and historical values of the variables (Howrey and Kelejian, 1971). Moreover, Wallis (1982) gives examples where the use of deterministic simulation predictors from a nonlinear econometric model may be less efficient than an *extrapolative time-series* forecast, in contrast with a well known result on the relative efficiency in the case of linear systems.

Replacing deterministic solution predictors with an estimate of the conditional means of endogenous variables in the prediction period allows to overcome this problem, to some extent.

Observing that deterministic solution of a nonlinear model with estimated parameters produces asymptotically biased and inefficient estimates of the conditional means, Mariano and Brown (1983) recommend to use the sample mean of replicated parametric stochastic simulations (using the algorithms proposed by Nagar, 1969, or by McCarthy, 1972, if the random errors are supposed multivariate normal), or (Brown and Mariano, 1984, 1985) to use the sample mean of nonparametric residual-based simulations. Both procedures attain, under certain conditions, asymptotic unbiasedness (and efficiency to some extent).

Empirical studies have been performed on a variety of real-world macroeconomic models used in practice by model builders for forecasting purposes (e.g. Bianchi et al, 1976, 1980, Calzolari, 1979, Fair, 1980, Hall, 1984, 1986, Fisher and Salmon, 1986). They aimed at evidencing whether considering the conditional mean of the endogenous variables in place of the deterministic solution gives or not significant improvements to the forecaster.



The conclusion is clearly model-dependent and not univocal; cases with significant differences have been evidenced.

Predictors obtained as conditional means of the endogenous variables are not, however, optimal from all points of view. They are obviously optimal if the forecaster has in mind some kind of quadratic loss function for his forecasts, but have two highly undesirable properties. The first is that

> *the means do not necessarily satisfy the equations of the model; given the values of the predetermined variables and of the model's parameters, it is generally impossible to find values of the random error terms which correspond to the mean values of all the endogenous variables.*

This is a problem in particular for those variables which appear in nonlinear identities; if the means are computed with replicated stochastic simulations (parametric or nonparametric), although all identities hold replication by replication, they do not hold anymore (at least the nonlinear ones) in terms of the means. In other words, the conditional mean does not provide a coherent forecast, and this can lead to some serious misinterpretations of the results. Suppose for example that an Italian model is designed to predict the exchange rate in lire per dollar ($ER1$). Then, in order to compute the exchange rate in dollars per lire, the model simply takes the reciprocal: the identity $ER2 = 1 / ER1$ will therefore be included among the equations. If we use the conditional means as predictors, the mean of $ER2$ is not the reciprocal of the mean of $ER1$. The naive analyst, who looks at the table of results supplied by the model, will probably think that these results are wrong (unless he is so ingenuous to believe that the model also considers the existence of arbitrage opportunities in the exchange market).

Considering the problem of incoherency of predictors obtained from conditional means, Hall (1986) observes that also the *modes* of the marginal distributions of the endogenous variables would not escape the problem: the *mode* of the product of two random variables is generally not the product of the two *modes*.

A second undesirable property of the *mean predictors* is also usually expected from the analysis of simple univariate distributions:

*the mean is usually expected to be less probable than the deterministic predictor; with respect to the deterministic solution value, the mean of an endogenous variable is shifted towards the side of more unlikely values, rather than towards the side where the probability density grows.*

In fact in skewed univariate distributions which are unimodal and of moderate asymmetry, an empirical relationship holds between *mean, median* and *mode* (see Kendall and Stuart, 1969, section 2.11), and it is that the three quantities occur on the distribution in the same order (or in the reverse order) as in the dictionary. This should imply that, if we associate the median of each endogenous variable to the deterministic solution of the model, the most likely values of the variable are systematically expected to lie on the opposite side of the mean.

What we propose in this paper is to reconsider the *mode of the joint distribution* of the endogenous variables as an optimal predictor which preserves identities. Although it does not coincide with the mode of the univariate distributions of the endogenous variables considered separately (i.e. their marginal distributions), in some sense it might be regarded as an estimate of *the most likely joint value* of all the endogenous variables simultaneously. At the same time, since it is obtained as solution of the system corresponding to a particular set of values of the random error terms, it implies coherency.

In a previous study, Brillet, Calzolari and Panattoni (1986) proposed to estimate the *mode of the joint distribution,* maximizing the joint density of a subset of the endogenous variables, corresponding to stochastic equations only. This was analogous to the approach that is usually followed for estimating models with *full information maximum likelihood;* in the first step of the estimation process, at least conceptually, identities are substituted into stochastic equations, and only after this the likelihood can be defined (and maximized) for the model that is, now, *without identities.*

In this paper, a more general approach is developed; it is explicitly designed to keep the identities. The model with identities is viewed as a mapping between the space of the random errors and an hypersurface in the higher dimensional space of the endogenous variables, and maximization is performed on such a hypersurface. This provides the *mode* of the joint distribution of *all* the endogenous variables.



Some technical difficulties, arising when searching for the maximum of the joint density function of all the endogenous variables in a large scale macroeconometric model, had to be solved. The computational problems, however, are only slightly more complex than for the method in Brillet et al. (1986).

Experimental results on the two different *mode predictors* are provided for a macro model of the Italian economy. For the sake of completeness and to allow comparisons, we also summarize in section 5 methods for computing the *mean predictors* and provide, also for them, experimental results.

## 2. NOTATIONS AND FIRST ASSUMPTIONS

Let the simultaneous equation model be represented as

$$f(y_t, x_t, a) = \begin{bmatrix} u_t \\ 0 \end{bmatrix} \qquad t = 1,2,\ldots,T \tag{1}$$

where $y_t$ is the $M \times 1$ vector of endogenous variables at time $t$, $x_t$ is the vector of predetermined variables at time $t$, $a$ is the vector of all unknown structural coefficients in the model, and $f$ is the $M \times 1$ vector of structural form operators. The model is supposed to contain $m \leq M$ stochastic equations and $M - m$ identities. The $m \times 1$ vector of random error terms at time $t$,

$$u_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{m,t} \end{bmatrix} \tag{2}$$

is assumed to be independently and identically distributed as $N(0, \Sigma)$ with the $m \times m$ covariance matrix $\Sigma$ completely unknown, apart from being symmetric and positive definite. The vector $u_t$ is followed by a vector of $M - m$ zeroes in the structural form equations (1). We also decompose the vector $y_t$ in two subvectors of length $m$ and $M - m$, respectively

$$y_t = \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{m,t} \\ y_{m+1,t} \\ \vdots \\ y_{M,t} \end{bmatrix}. \tag{3}$$

There is a certain freedom in ordering the endogenous variables inside the subvectors $y_t^{(1)}$ and $y_t^{(2)}$. A choice that is rather obvious, but not necessarily unique, is to put into $y_{1,t}$ the variable which is *explained* by the structural equation whose error term is $u_{1,t}$, put into $y_{2,t}$ the variable which is *explained* by the structural equation whose error term is $u_{2,t}$, and so on, till $y_{m,t}$ and $u_{m,t}$. We have filled in this way the subvector $y_t^{(1)}$. The subvector $y_t^{(2)}$ is filled in some way from the remaining $M - m$ endogenous variables.

It is usually assumed that a simultaneous equations system like (1) uniquely defines the values of the elements of $y_t$ once values for the coefficients, the predetermined variables, and the disturbance terms are given. This means that the structural form equations (1) implicitly define a system of reduced form equations

$$y_t = g(x_t, \alpha, u_t) \tag{4}$$

Since all predictors are *conditional* on model's parameters and predetermined variables, we shall often indicate the reduced form simply as

$$y_t = g(u_t) \tag{5}$$



## 3. THE MODE PREDICTORS

### 3.1. A simplified approach

The way the problem is tackled in this section reminds the *full information maximum likelihood* estimation of simultaneous equations systems (e.g. Amemiya, 1983). Since the model (1) includes $M - m$ identities, we first substitute these identities into the $m$ stochastic equations.[2] The resulting model may be written as

$$p(y_t^{(1)}, x_t, \alpha) = u_t \qquad t = 1,2,\ldots,T \tag{6}$$

where the $m \times 1$ vector of functions $p$ is obtained from the vector of functions $f$ (eq.1) after substitutions.

We assume that, given the coefficients and the predetermined variables, the functions vector $p$ such that $u_t = p(y_t^{(1)})$ is a continuous and differentiable *one to one* mapping from a subset of $R^m$ *onto* the whole $R^m$ ($u_t$ is multivariate normal, and therefore it spans the whole $R^m$ space, while $y_t^{(1)}$ may be restricted to a subspace: for example, some of its elements may not assume negative values) and that the inverse function is also continuous and differentiable.

The joint density of the elements of $y_t^{(1)}$ can thus be obtained, as usual, from the density of $u_t$ and the Jacobian determinant. Apart from an additive constant, the log-density of the $m$ - dimensional random vector $y_t^{(1)}$ is

[2] The notations could be made more accurate by dividing the vector $f$ in two subvectors of functions, and considering more carefully the mappings implied by them. A more precise notation, however, is not strictly necessary in what follows, and is not introduced for the sake of simplicity.

$$(7) \qquad L_t = -\frac{1}{2}\log|\Sigma| + \log\left|\left|\frac{\partial p_t}{\partial y_t^{(1)\prime}}\right|\right| - \frac{1}{2}p_t'\,\Sigma^{-1}p_t$$

where the vertical bar indicates *absolute value*, while the slash indicates *determinant*.

To obtain the *mode predictor* at time $h$, we must first compute the subvector $y_h^{(1)}$ that maximizes (7), given the predetermined variables $x_h$, the coefficients $a$ and the covariance matrix $\Sigma$ (for $a$ and $\Sigma$, of course, we shall use the available estimates $\hat{a}$ and $\hat{\Sigma}$). Given $y_h^{(1)}$, we must then solve the subsystem of identities, obtaining $y_h^{(2)}$, and thus the complete $M \times 1$ predictor for all the endogenous variables.

It is well known from the theory of maximum likelihood estimation that the boring operation of substituting the identities can be bypassed (see Oberhofer, 1971, who extends to the nonlinear case the treatment of identities in maximum likelihood estimation given by Rothenberg and Leenders, 1964, pp.71-72; see also Anderson, 1984, theorem A.3.2). In fact, the same values (and therefore the same maximum) of $L_t$ can be obtained if we partition the $M \times M$ Jacobian corresponding to all equations into 4 blocks

$$(8) \qquad J = \frac{\partial f_h}{\partial y_h'} = \begin{bmatrix} J_{1,1} & J_{1,2} \\ J_{2,1} & J_{2,2} \end{bmatrix}$$

(the 1,1 block corresponds to stochastic equations) and instead of $|\partial p_h/\partial y_h^{(1)\prime}|$ we use in (7) the ratio between the determinant of the $M \times M$ Jacobian matrix of the complete system, and the determinant of the $(M - m) \times (M - m)$ submatrix corresponding to identities

$$(9) \qquad L_t = -\frac{1}{2}\log|\Sigma| + \log\left|\frac{|J|}{|J_{2,2}|}\right| - \frac{1}{2}p_t'\,\Sigma^{-1}p_t$$



To simplify notations, in what follows the *time subscripts* $t$ and $h$ will sometimes be suppressed, when this does not create confusion.

For an easier comparison with the approach of next section, we modify the expression of the density given in equation (7). First, rather than viewing $L$ as a function of $y^{(1)}$, we view it as a function of $u$. To do this, we simply substitute $u$ to $p$ in the last term, and consider the Jacobian matrix of first derivatives of the elements of $y^{(1)}$ with respect to the elements of $u$ (which is simply the inverse of the matrix $\partial p/\partial y^{(1)\prime}$ used above)

$$(10) \qquad -\frac{1}{2}\log|\Sigma| - \log\left|\left|\frac{\partial y^{(1)}}{\partial u'}\right|\right| - \frac{1}{2}u'\,\Sigma^{-1}u$$

Then we simply substitute to the absolute value of the Jacobian determinant the square root of the determinant of the product between the transpose of the Jacobian matrix and the matrix itself

$$(11) \qquad -\frac{1}{2}\log|\Sigma| - \log\left|\frac{\partial y^{(1)\prime}}{\partial u}\,\frac{\partial y^{(1)}}{\partial u'}\right|^{\frac{1}{2}} - \frac{1}{2}u'\,\Sigma^{-1}u$$

Maximizing (11) in the forecast period, $h$

$$(12) \qquad \underset{u \in R^m}{Max}\left[-\frac{1}{2}\log|\Sigma| - \log\left|\frac{\partial y^{(1)\prime}}{\partial u}\,\frac{\partial y^{(1)}}{\partial u'}\right|^{\frac{1}{2}} - \frac{1}{2}u'\,\Sigma^{-1}u\right]$$

provides a value for the $m \times 1$ vector of random errors $u_h$. Inserting this vector into the model (1) and solving the model at time $h$ (that is, at least conceptually, the same as inserting it into the reduced form equations 4 or 5), we obviously get the same $M \times 1$ predictor $y_h$ discussed above.

The values of the endogenous variables computed in this way provide only a partial answer to our problem. In fact, the value computed for the first subvector ($y_h^{(1)}$) is the mode of the joint distribution of the elements of this

subvector, but the second subvector ($y_\lambda^{(2)}$) is simply calculated to preserve the internal coherency of the model (the whole vector $y_\lambda$ is, in fact, a solution of the system). Suppose now that we interpret the *mode predictor* as the *most likely joint value* of the endogenous variables in the forecast period. Then only $y_\lambda^{(1)}$ can be interpreted in this way, but the whole vector $y_\lambda$ is *not* the *most likely joint value* of *all* the endogenous variables.

### 3.2. A more general approach

In order to maximize the joint density function of *all* the endogenous variables of the model, we must first consider the type of mapping implied by the model, then some concepts on the measure of a regular (hyper)surface, and finally introduce a suitable definition of the probability density, with respect to this measure. To clarify the problem, let us consider the simplest cases.

Suppose that our model consists of two equations, the former is stochastic while the latter is an identity. Given coefficients and predetermined variables, solving the model we get the value of the *two* endogenous variables for *one* value of the error term. If the solution is unique and the error term is normally distributed, this gives a mapping from $\mathbb{R}^1$ *into* $\mathbb{R}^2$. The reverse is not true; we can, in fact, enter $y_{1,t}$ and $y_{2,t}$ into the first equation and compute $u_{1,t}$ as a *residual* (and this will certainly be unique), but we cannot take two arbitrary values for the endogenous variables, since they are constrained by the second equation (a point that is well known to the model builders: coherent historical data *must* satisfy all the identities). The values of $y_{1,t}$ and $y_{2,t}$ must lie on a curve in the two-dimensional plane. Therefore our model can be viewed as a mapping from $\mathbb{R}^1$ *onto* a one-dimensional subset of $\mathbb{R}^2$, that is a curve in the plane.

Suppose that our model consists of three equations, two of which are stochastic. If for any value of the random errors $u_{1,t}$ and $u_{2,t}$ (in $\mathbb{R}^2$) the model provides a unique solution for $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$, the values of the $y$'s will neither span the full $\mathbb{R}^3$ space, nor any three-dimensional subset of $\mathbb{R}^3$, but



will lie on a surface. The model, therefore, can be regarded as a mapping from $\mathbb{R}^2$ *onto* a two-dimensional surface in $\mathbb{R}^3$.

In the general case of $M$ equations, $m$ of which are stochastic, the model (in the forecast period, given predetermined variables and parameters) can be viewed as a mapping from $\mathbb{R}^m$ *onto* an $m$-dimensional hypersurface in the $\mathbb{R}^M$ space (being $m \leq M$). We call $Y$ this hypersurface. Under the two assumptions

1) for any $u$ in $\mathbb{R}^m$ the solution $y$ is unique;

2) for any $y$ in $Y$ we get a unique vector $u$;

then we are dealing with a *one to one* mapping between $\mathbb{R}^m$ and the $m$-dimensional hypersurface $Y$ in $\mathbb{R}^M$. As in equation (5), we indicate this mapping as $g: \mathbb{R}^m \to Y$ (the reduced form) and its inverse as $g^{-1}: Y \to \mathbb{R}^m$ (practically the first $m \times 1$ subvector of the structural form functions vector $f$).

The first assumption is somewhat restrictive, but widely adopted and accepted in the treatment of nonlinear systems. The second, on the contrary, is not restrictive at all, given the way in which model builders write the structural form equations of a model like (1). In fact, if a vector of endogenous variables belongs to $Y$, it certainly satisfies the constraint of the $M - m$ identities. In such a case, if we plug this vector into the $m$ stochastic equations, we get a unique vector of *residuals*.

Having defined the nature of this *one to one* mapping, we now assume, as in the previous section, continuity and differentiability (in both directions), and finally assume that the $M \times m$ Jacobian matrix $G(u) = \partial y/\partial u' = \partial g(u)/\partial u'$ has full rank ($= m$).

In a $M \times m$ rectangular matrix, a substitute for the notion of determinant of a square matrix is the concept of *modulus*. It is defined as the square root of the sum of all the $\binom{M}{m}$ squared determinants of order $m$ obtained by cancelling $M - m$ rows. A simple way of computing the modulus, even when $M$ and $m$ are large, is obtained applying a well known theorem of linear algebra on the determinant of the product of two rectangular matrices (e.g.

Hadley, 1961, p.102); it is the square root of the determinant of the product between the $m \times M$ transpose and the matrix itself. For our Jacobian, $G(u)$, it is

$$(13) \qquad |G(u)| = \left| \frac{\partial y}{\partial u'} \right| = \left| \frac{\partial g(u)}{\partial u'} \right| = \left/ \left[ \frac{\partial y'}{\partial u} \frac{\partial y}{\partial u'} \right] \right.^{\frac{1}{2}}$$

Having assumed full rank ($= m$) for the Jacobian, the modulus is $\neq 0$. The $m$- dimensional hypersurface $Y$ defined by the model is, under all these assumptions, a *regular hypersurface.*

We need now to define a measure on the regular $m$- dimensional hypersurface $Y$. As well known, the $M$-dimensional Lebesgue measure of any $m$- dimensional hypersurface in $\mathbb{R}^M$ ($m < M$) is zero (e.g. Sikorski, 1969, p.299). This is consistent with intuition: the two-dimensional measure (area) of a curve in the plane is zero; the three-dimensional measure (volume) of a curve or of a surface in $\mathbb{R}^3$ is also zero, etc. But we can define a special one-dimensional measure on curves, consistent with the intuitive notion of length; or a special two-dimensional measure on surfaces, consistent with the intuitive notion of area. Existence and uniqueness of such a measure, for a regular $m$-dimensional hypersurface $Y$ in $\mathbb{R}^M$ ($m \leq M$) is stated, for example, in Sikorski (1969, p.327, theorem 1.1), and is given by

$$(14) \qquad d\mu_Y = |G(u)|\, du = \left| \frac{\partial y}{\partial u'} \right| du = \left| \frac{\partial g(u)}{\partial u'} \right| du$$

This theorem can be applied to derive a probability density with respect to the hypersurface measure $\mu_Y$. If $\varphi(u)$ is the probability density for $u$ in $\mathbb{R}^m$, the measure of the probability that $y$ belongs to an element of the hypersurface $Y$, resulting (through $y = g(u)$) from the element $du$ of $\mathbb{R}^m$, is given by



$$(15) \qquad d\pi_Y = \varphi(u)\, du$$

and can easily be developed as follows

$$(16) \qquad d\pi_Y = \varphi(u)\, du = \frac{\varphi(u)}{|G(u)|} |G(u)|\, du$$

$$= \frac{\varphi\{g^{-1}[y]\}}{|G\{g^{-1}[y]\}|} d\mu_Y = \psi_Y(y)\, d\mu_Y$$

where

$$(17) \qquad \psi_Y(y) = \frac{\varphi\{g^{-1}[y]\}}{|G\{g^{-1}[y]\}|}$$

is taken as the definition of the probability density of the vector $y$ on the regular $m$-dimensional hypersurface $Y$. It is rather obvious that $\psi_Y(y)$ respects the usual conditions for density functions, being always $\geq 0$ and being equal 1 its integral, over the whole $Y$, with respect to the $m$-dimensional hypersurface measure $\mu_Y$.

We now indicate with $L_Y$ the log-density

$$(18) \qquad L_Y = \log[\psi_Y(y)]$$

that will obviously be $L_{Y,t}$ for the values of $y$ at time $t$, or $L_{Y,h}$ in the forecast period, $h$. Being $\varphi$ the density of the $m$-variate normal $N(0, \Sigma)$, we have, apart from a constant term

$$(19) \qquad L_Y = -\frac{1}{2}\log|\Sigma| - \log\left|G\left\{g^{-1}[y]\right\}\right| - \frac{1}{2}\left[g^{-1}(y)\right]'\Sigma^{-1}\left[g^{-1}(y)\right]$$

$L_Y$ in equation (19) is *conceptually* defined as a function of the $M \times 1$ vector $y$, whose domain is the $m$ - dimensional hypersurface $Y$ in $\mathbb{R}^M$. The $M \times 1$ vector $y$, that maximizes $L_Y$ in $Y$ (that is, $y$ is *constrained* on $Y$), is the *mode predictor*

$$(20) \quad \underset{y \in Y}{Max} \left\{ -\frac{1}{2}\log|\Sigma| - \log\left|G\left\{g^{-1}[y]\right\}\right| - \frac{1}{2}\left[g^{-1}(y)\right]'\Sigma^{-1}\left[g^{-1}(y)\right] \right\}$$

If we interpret it as the *most likely joint value* of the endogenous variables in the forecast period, such an interpretation applies to *all* the endogenous variables, unlike the result derived in the previous section. Numerical values of $L_Y$ can easily be calculated by replacing $g^{-1}(y)$ with $u$

$$(21) \qquad -\frac{1}{2}\log|\Sigma| - \log|G(u)| - \frac{1}{2}u'\Sigma^{-1}u$$

Equation (21) provides a function of the $m \times 1$ vector $u$, that must be maximized for $u$ *without constraints* in $\mathbb{R}^m$

$$(22) \qquad \underset{u \in \mathbb{R}^m}{Max} \left[ -\frac{1}{2}\log|\Sigma| - \log|G(u)| - \frac{1}{2}u'\Sigma^{-1}u \right]$$

in the forecast period, $h$. Inserting the resulting vector $u_h$ into the model (1), then solving the model in the forecast period, we obtain the $M \times 1$ *mode predictor* $y_h$.

From a technical point of view, the process is quite similar to what we did in the previous section (eq.12). The only difference is that in equation (12) we have the modulus of the $m \times m$ square matrix $\partial y^{(1)}/\partial u'$, while here we have the modulus of the $M \times m$ rectangular matrix $\partial y/\partial u'$.



## 4. MAXIMIZATION ALGORITHM

The two functions to be maximized for the methods of sections 3.1 and 3.2 are given in equations (11) and (21), respectively. Maximization is to be done with respect to the vector $u_h$, while predetermined variables are given, as well as coefficients and covariance matrix of the random error process (set at their estimated values $\hat{a}$ and $\hat{\Sigma}$). The first term in both equations is a constant. We must therefore compute the vector $u_h$ which maximizes the sum of the second and third term of equations (11) and (21). The third term in both equations is particularly simple, being a positive definite quadratic form, whose Hessian is the matrix $\Sigma^{-1}$.

We have used the well known updating formula due to Broyden, Fletcher, Goldfarb and Shanno (BFGS, see for example Dennis and More', 1977). The algorithm is based on an iterative updating of an initial ($m \times m$) positive definite matrix. Since the computational efficiency is greatly improved if the initial matrix approximates the Hessian, it was rather obvious in both cases to use the available estimate of $\Sigma^{-1}$.

The algorithm also requires the evaluation at each step of the gradient of the function. This revealed to be a rather serious problem and at this stage it has been solved by the numerical computation of first derivatives. This approach however has two main drawbacks. First of all it requires a long computation time, but nevertheless it came out to be computationally more efficient than rival maximizing algorithms which require only the computation of the function value. To quantify this aspect we can mention that in the case of the Italian model here considered (see below for description) the maximization of the quantities defined by equations (11) and (21), with a tolerance of ten significant digits on the function's value, in both cases required five iterations. The computation globally took about 14 minutes of CPU time on an IBM 4341 computer in both cases. However for the method of section 3.1 we can maximize the expression of the density given in equation (9), rather than (11); in this case the computation is much faster (only 50 seconds). This is mainly due to the fact that the elements of

the Jacobian matrices in equations (11) and (21) are obtained from differences between two solutions of the simultaneous equation system, while the corresponding elements in equation (9) are simply computed from the differences between the residuals of the single equations: obviously the latter computation is much faster than the former. Unfortunately, we are unable to provide a similar computational simplification for the method of section 3.2.

A second main drawback lies in a possible lack of accuracy in the computation of the derivatives. In order to assess the robustness of the results versus the way in which derivatives are computed, several formulas (i.e. two, three and five points formulas with different sizes of the increment) were experimented with. The results proved to be very robust versus both the choice of the formula and the choice of the numerical increment, provided that a centered formula is used.

## 5. THE MEAN PREDICTORS

We use the reduced form notation given in equation (4). [3] If the vector of functional operators $g$ is representable in closed form, and the analytical computation of the conditional mean is feasible

$$(23) \qquad \gamma(x_t, a, \Sigma) = E(y_t \mid x_t, a, \Sigma) = E[g(x_t, a, u_t) \mid x_t, a, \Sigma]$$

then if $\hat{a}$ and $\hat{\Sigma}$ are the available estimates of the structural form parameters, the estimated conditional mean in the forecast period $h$, $\gamma(x_h, \hat{a}, \hat{\Sigma})$ may be

[3] Most of this section is taken from the unpublished paper by Brillet, Calzolari and Panattoni (1986). It is reported here for the sake of completeness, to allow the reader an easier understanding of the computations underlying the numerical results of the last section.



used as predictor. Nonlinearity usually implies this predictor to be different from the usually employed deterministic predictor $g(x_h, \hat{a}, 0)$.

Since, however, the vector of functional operators $g$ is generally not representable in closed form in the case of nonlinear models, and the analytical computation of the mean (the vector $\gamma$) is usually infeasible, some approximation techniques must be employed.

The approximation technique that is most widely used in the literature (e.g. Bianchi et al., 1976, Fair, 1980, Fisher and Salmon, 1986, Hall, 1986) is the stochastic simulation procedure with normal errors. The procedure is as follows.

1) A vector of pseudo-random numbers $\tilde{u}_h$, with multivariate normal distribution, zero mean and the available covariance matrix $\hat{\Sigma}$, is generated. The method of Nagar (1969) can be applied if $\hat{\Sigma}$ is positive definite; if $\hat{\Sigma}$ is not of full rank, the method of McCarthy (1972) can be used.

2) The vectors $\tilde{u}_h$ are inserted into the model, where the structural coefficients are maintained fixed at their originally estimated values, and the model is solved in the forecast period, $h$, obtaining for the endogenous variables the vector $\tilde{y}_h$.

Stages 1 and 2 are repeated and sample means of the elements of $\tilde{y}_h$ are computed.

If finite moments exist, a very large number of replications would lead, in practice, to the exact values of the means $\gamma(x_t, a, \Sigma)$, if the parameters of the model (the vector $a$ and the covariance matrix of the structural disturbances) were known with certainty. As, however, we assume only estimates of these parameters, stochastic simulation will lead to an estimate of the conditional means of the endogenous variables in the prediction period $\gamma(x_h, \hat{a}, \hat{\Sigma})$.

The experimental variance of the sample mean decreases in inverse proportion with the number of replications. This is often insufficient to allow appreciating significant differences between the mean and the deterministic

solution value even with a rather large number of replications (Bianchi et al., 1976, 1980, Fisher and Salmon, 1986). Huge gains in the computational efficiency are often obtained in macroeconometric models by the use of *antithetic variates* in the stochastic simulation procedure (Calzolari, 1979, Brown and Mariano, 1985). Replications are performed pairwise, once with pseudo-random error terms $\tilde{u}_h$ generated as discussed above, and once with the same vector of error terms with opposite sign, $-\tilde{u}_h$. The results of the two replications (which will presumably be negatively correlated) are averaged, and the means of the endogenous variables are computed as sample means of the pairwise means.

Whether or not *antithetic variates* are used, the results do not change, except that the same accuracy may be obtained (and has been obtained for the model we are considering in this paper) with a smaller number of replications. The results we present have been obtained with 50000 couples of replications, which guaranteed for all the variables displayed in the tables an estimate of the bias (deterministic solution minus conditional mean) at least 50 times larger than its experimental standard deviation: a similar computational accuracy without *antithetic variates* would be obtained only at the cost of several millions of independent replications. [6]

[6] Alternative *mean predictors* could be obtained with the *residual-based* procedures proposed by Brown and Mariano (1984, 1985). These procedures utilize complete enumeration of the residuals over the sample period. The basic procedure consists of replicating the solution of the model in the forecast period, $h$, exactly $T$ times, using the $T$ vectors of estimated residuals $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_T$ and then computing sample means of the elements of the $T$ vectors of solutions. The computational stages are quite similar to those of the parametric method described above. In the first stage, however, we use one vector of estimated residuals rather than using a generator of pseudo-random numbers; moreover, stages 1 and 2 are repeated *exactly* $T$ times (rather than an arbitrary number of times), using each time a different vector of residuals. Since this paper is focussed on the *mode predictors*, results on the *mean predictors* are given only for comparison purposes and are therefore confined to one method only.



## 6. A MACRO MODEL OF THE ITALIAN ECONOMY

In a recent paper, Sylos Labini (1987) has analyzed the main characteristics of the labor market in Italy. Some structural equations are also presented in the paper, for unemployment, productivity, wages and salaries, sometimes disaggregated by sectors. These equations have then been plugged into a macroeconometric model of the real sector (Damiani, 1987).

The present version of the model consists of six blocks: demand, supply, prices, labor market, foreign trade, income distribution. The total number of equations in the model is $M = 62$, while $m = 25$ is the number of stochastic equations.

Estimates of the structural coefficients of the model have been obtained by means of a straightforward extension of Brundy and Jorgenson's (1971) instrumental variables method (limited information) to the case of nonlinear models. The method was applied iteratively, as proposed in Dutta and Lyttkens (1974), so that the final estimates of coefficients are not affected by the choice of the initial values of coefficients. In each iteration, the instrumental variables are computed with deterministic solution of the system (which is the simplest choice, although not the *best* in the class of nonlinear limited information estimators, as is explained in Amemiya, 1983).

The last year of the sample period is 1985. Forecasts are related to the first year outside the sample estimation period, 1986.

## 7. RESULTS ON THE ITALIAN MODEL

For some of the main endogenous variables of the model (listed in table 1), table 2 presents in the first column, indicated as *mean*, the differences between the deterministic predictor and the *mean predictor* computed as in

section 5. For each variable, this difference is displayed in *per thousand* form

$$(24) \qquad \frac{y_{i,h}^{(det)} - y_{i,h}^{(mean)}}{y_{i,h}^{(det)}} \cdot 1000$$

The last two columns of the table, labelled *M-mode* and *m-mode*, display, still in *per thousand* form, the differences between the deterministic predictor and the two *mode predictors* defined in sections 3.2 and 3.1, respectively.

$$(25) \qquad \frac{y_{i,h}^{(det)} - y_{i,h}^{(mode)}}{y_{i,h}^{(det)}} 1000$$

As far as the *mean predictor* is concerned, the difference with respect to the deterministic predictor can be indicated as *bias*. The size of the bias is rather small for almost all variables, although we must consider that we are dealing only with forecasts *one year ahead*. For instance, it is -4.6 ‰ for private consumptions (*COFAMK_I*), but it is only 0.49 ‰ for the growth rate of *GDP* (*PILK$I*).

It has been shown (see Kendall and Stuart, 1969, section 2.11) that in most univariate cases the difference between *mean* and *mode* is about three times the difference between *mean* and *median* (which can be assimilated to the deterministic value). This is equivalent to saying that the second statistic (*median - mode*) should be minus two times the first one (*median - mean*). This is obviously not true in our case, whether we consider the *mode* computed with the simplified approach of section 3.1, or the global method of section 3.2 (fourth and third columns of table 2, respectively). Therefore, abandoning the deterministic forecast in favor of the (more efficient) *mean predictor*, we are not necessarily moving towards a region of less likely values.

Between the two *mode predictors*, the global approach of section 3.2 produces values that are considerably more distant from the deterministic



solution. The distance is also considerably larger than for the *mean predictors*.

As a final remark we must point out that, on the basis of other experiments carried out with different non linear models (like the Klein-Goldberger model), this behaviour resulted to be substantially model independent.

**Table 1**

List of the main endogenous variables

| | |
|---|---|
| *PREMIN$I* | = Deflator for domestic private consumption (growth rate). |
| *PREXP$I* | = Export price index (growth rate). |
| *P2$I* | = Production per hour in the industrial sector (growth rate). |
| *S$I* | = Wage per hour in the industrial sector (growth rate). |
| *INVIDK$I* | = Gross fixed investment in the industrial sector (growth rate). |
| *PR7* | = Profit rate. |
| *PREING$I* | = Wholesale price index (growth rate). |
| *VAINDK$I* | = Value added in the industrial sector at constant prices (growth rate). |
| *DIST$I* | = Unemployment (growth rate). |
| *OCCD$I* | = Total dependent employment (growth rate). |
| *OCSRD$I* | = Employment in the sector of services (growth rate). |
| *EXFOBK$I* | = Exports at constant prices (growth rate). |
| *IMFOBK$I* | = Imports at constant prices (growth rate). |
| *PILK$I* | = Gross domestic product at constant prices (growth rate). |
| *COFAMK_I* | = Domestic private consumptions. |

Table 2

One-period forecast at 1986
Per thousand deviations from deterministic forecasts

| | mean | M - mode | m - mode |
|---|---|---|---|
| PREMINSI | -1.262 | -.3975 | .0445 |
| PREXPSI | -3.943 | 9.592 | .6288 |
| PSI | .2028 | -33.57 | -4.008 |
| SSI | -3.382 | 1.271 | .0216 |
| INVIDKSI | .7541 | 8.938 | -1.401 |
| PRT | .2050 | -.8109 | -.1591 |
| PREINCSI | -5.935 | 2.931 | .0883 |
| YAINDKSI | .8387 | -10.79 | -2.864 |
| DISTSI | .4952 | 54.41 | 2.169 |
| OCCDSI | .4054 | 26.64 | -.3355 |
| OCSRDSI | .0562 | 1.788 | -.0155 |
| EXFOBKSI | 1.485 | 1.308 | .3365 |
| IMFOBKSI | -3.897 | -15.22 | -2.805 |
| PILKSI | .4982 | 5.811 | -1.721 |
| COFAMK_I | -4.869 | 11.49 | -2.439 |



## REFERENCES


Amemiya, T. (1983): "Non-Linear Regression Models", in *Handbook of Econometrics*, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland, Vol. I, 333-389.

Anderson, T.W. (1958): *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley.

Bianchi, C., and G. Calzolari (1980). "The One-Period Forecast Errors in Nonlinear Econometric Models", *International Economic Review* 21, 201-208.

Bianchi, C., G. Calzolari, and P. Corsi (1976): "Divergences in the Results of Stochastic and Deterministic Simulation of an Italian Non-Linear Econometric Model", in *Simulation of Systems*, ed. by L. Dekker. Amsterdam: North Holland, 653-661.

Brillet, J.L., G. Calzolari, and L. Passaglia (1986): "Coherent Optimal Prediction with Large Nonlinear Systems: An Example Based on a French Model". Paris: INSEE, Service des Programmes, discussion paper presented at the *European Meeting of the Econometric Society*, Budapest, September 1-5.

Brown, B.W., and R.S. Mariano (1984): "Residual-Based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System", *Econometrica* 52, 321-343.

Brown, B.W., and R.S. Mariano (1985): "Reduced Variance Prediction in Nonlinear Simultaneous Systems". Rice University, discussion paper presented at the *World Congress of the Econometric Society*, M.I.T., Cambridge, MA, August 17-24.

Brundy,J.M., and D.W.Jorgenson (1971): "Efficient Estimation of Simultaneous Equations by Instrumental Variables", *The Review of Economics and Statistics* 53, 207-224.

Calzolari,G. (1979): "Antithetic Variates to Estimate the Simulation Bias in Non-Linear Models", *Economics Letters* 4, 323-328.

Damiani,M. (1987): "Uno Schema Macroeconomico di Simulazione per l'Economia Italiana". Universita' di Perugia, Facolta' di Scienze Politiche, mimeo.

Dennis,J.E., and J.J. More' (1977): "Quasi-Newton Methods, Motivation and Theory", *SIAM Review* 19, 46-89.

Dutta,M., and E.Lyttkens (1974): "Iterative Instrumental Variables Method and Estimation of a Large Simultaneous System", *Journal of the American Statistical Association* 69, 977-986.

Fair,R.C. (1980): "Estimating the Expected Predictive Accuracy of Econometric Models", *International Economic Review* 21, 355-378.

Fisher,P., and M.Salmon (1986): "On Evaluating the Importance of Nonlinearity in Large Macroeconometric Models", *International Economic Review* 27, 625-646.

Hadley,G. (1961): *Linear Algebra*. Reading, MA: Addison - Wesley.

Hall,S.G. (1984): "The Application of Stochastic Simulation Techniques to the National Institute's Model 7". London: NIESR, discussion paper No.65.

Hall,S.G. (1986): "The Importance of Non-Lincarities in Large Forecasting Models with Stochastic Error Processes", *Journal of Forecasting* 5, 205-215.



Howrey,E.P., and H.H.Kelejian (1971): "Simulation versus Analytical Solutions: the Case of Econometric Models", in *Computer Simulation Experiments with Models of Economic Systems*, ed. by T.H.Naylor. New York: John Wiley, 299-319.

Kendall,M.G., and A.Stuart (1969): *The Advanced Theory of Statistics*, Vol.I. London: Charles Griffin.

Mariano,R.S., and B.W.Brown (1983): "Asymptotic Behavior of Predictors in a Nonlinear Simultaneous System", *International Economic Review* 24, 523-536.

McCarthy,M.D. (1972): "Some Notes on the Generation of Pseudo-Structural Errors for Use in Stochastic Simulation Studies", in *Econometric Models of Cyclical Behavior*, ed. by B.G.Hickman. New York: NBER, Studies in Income and Wealth No.36, 185-191.

Nagar,A.L. (1969): "Stochastic Simulation of the Brookings Econometric Model", in *The Brookings Model: Some Further Results*, ed. by J.S.Duesenberry, G.Fromm, L.R.Klein and E.Kuh. Amsterdam: North Holland, 425-456.

Oberhofer,W. (1971): "On the Treatment of Definition Equations in Likelihood Functions". Universität Bonn: Institut für Gesellschafts- und Wirtschaftswissenschaften, Wirtschaftstheoretische Abteilung, discussion paper No.14.

Rothenberg,T.J., and C.T.Leenders (1964): "Efficient Estimation of Simultaneous Equation Systems", *Econometrica* 32, 57-76.

Sikorski,R. (1969): *Advanced Calculus. Functions of Several Variables*. Warszawa: PWN - Polish Scientific Publishers.

Sylos Labini,P. (1987): "The Theory of Unemployment, too, is Historically Conditioned", *Banca Nazionale del Lavoro Quarterly Review* 163, 379-435.

Wallis,K.F. (1982): "Time Series versus Econometric Forecasts: A Non-Linear Regression Counter-Example", *Economics Letters* 10, 309-315.